Rethinking the Joint Optimization in Video Coding for Machines: A Case Study

Changsheng Gao, Zhuoyuan Li, Li Li, Dong Liu, and Feng Wu
Intelligent Visual Data Coding Lab, University of Science and Technology of China
{changshenggao, lil1, dongeliu, fengwu}@ustc.edu.cn
zhuoyuanli@mail.ustc.edu.cn

2024 Data Compression Conference (DCC) | 979-8-3503-8587-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/DCC58796.2024.00073

In this work, we investigate the joint optimization strategy in the scenario of video coding for machines (VCM). We formulated two kinds of joint optimization strategies, **Opt_JA** and **Opt_JH**, and compared them with the separate optimization strategy **Opt_S**. The three optimization strategies are illustrated in Fig. 1. In **Opt_S**, we separately train the feature compression network with mean squared error (MSE). In **Opt_JA**, we optimize all modules jointly toward the person re-identification task. In **Opt_JH**, only the aggregation module and feature compression module are jointly optimized. The feature compression consists of two fully-connected (FC) layers and two batch normalization (BN) layers. Specifically, we set five compression ratios (CR): 256, 128, 64, 32, and 16.

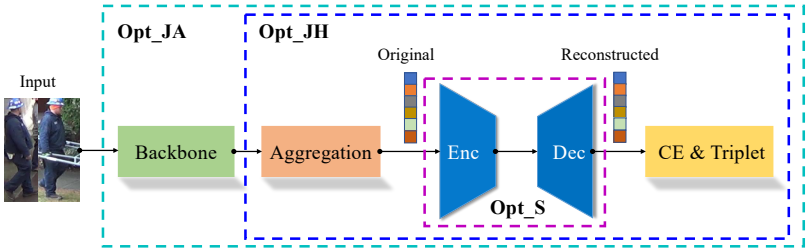

Figure 1: Joint optimization framework of person ReID and feature compression.

We verified the efficiency of the three optimization strategies on the DukeMTMC-reID dataset, and the results are presented in Table 1. At low bitrates, joint optimization is better than separate optimization, while separate optimization is better at high bitrates. **Opt_JA_O** and **Opt_JH_O** denote the performance on the original features.

Table 1: Performance comparison between different optimization strategies

CR	Opt_S	Opt_JH	Opt_JA	Opt_JH_O	Opt_JA_O
256	26.54	35.33	40.29	81.34	54.36
128	57.44	56.51	56.16	81.29	58.15
64	73.48	68.57	65.89	80.96	59.04
32	79.05	74.74	71.89	80.83	61.58
16	80.50	76.37	72.41	80.8	61.61

2375-0359/24/$31.00 ©2024 IEEE
DOI 10.1109/DCC58796.2024.00073